

Chu Xin (Cloris) Cheng

cheng2@caltech.edu | (805) 755-9772 | clorischeng.github.io

EDUCATION

California Institute of Technology

B.S. Computer Science

Pasadena, CA

Sep. 2021 – Jun. 2025

- **Current GPA:** 4.3
- **Relevant Coursework:** Data Mining and Machine Learning, Learning Systems, Statistical Inference, Fundamentals of Statistical Learning, Advanced Topics in Machine Learning, Applied Linear Algebra, Applied Real and Functional Analysis, Introductory Methods of Computational Mathematics, Introduction to Probability Models, Stochastic Processes and Regression, Mathematical Optimization.

LIST OF PUBLICATIONS

Practical Bayesian Algorithm Execution via Posterior Sampling

Chu Xin Cheng*, Raul Astudillo*, Thomas Desautels, Yisong Yue

Neural Information Processing Systems (NeurIPS), 2024.

Improving Sample Efficiency of High Dimensional Bayesian Optimization with MCMC

Zeji Yi*, Yunyue Wei*, Chu Xin Cheng*, Kaibo He, Yanan Sui

Learning for Dynamics & Control Conference (L4DC), 2024. (Link)

Preferential Bayesian Optimization with Multiple Mixed Objectives

Raul Astudillo, Kejun Li, Maegan Tucker, Chu Xin Cheng, Aaron Ames, Yisong Yue

Preprint. (Link)

HONORS AND AWARDS

Jack E. Froehlich Memorial Award

2024

California Institute of Technology

- The award is presented to a junior in the upper 5 percent of their class who shows outstanding promise for a creative professional career.

Henry Ford II Scholar Award

2024

California Institute of Technology

- The award is presented annually to engineering students with the best academic record at the end of the third year of undergraduate study.

WORK AND RESEARCH EXPERIENCES

Machine Learning Researcher, Yisong Yue's Group

2022 - 2024

California Institute of Technology / Lawrence Livermore National Lab

Project 1, 2022 Generalized traditional Bayesian optimization methods to learn from preference feedback within the dueling bandits framework, with an extension to continuous space utilizing Gaussian processes. Developed an algorithm that incorporates Langevin dynamics and Kalman filtering on the posterior distribution to leverage gradient feedback.

Project 2, 2023 Contributed to developing the first theoretical framework for multi-attributed preferential Bayesian optimization, aimed at its application to exoskeleton gait optimization. Designed an algorithm that employs Chebyshev scalarization to produce Pareto-optimal actions. **Paper:** Preferential Bayesian Optimization with Multiple Mixed Objectives.

Project 3, 2023 Investigated machine learning methods for protein engineering, focusing on antibody design. Applied conformal prediction to Graph Neural Networks to obtain confidence intervals with guaranteed coverage for modeling the fitness function with uncertainty.

Project 4, 2024 Developed a new Bayesian computation algorithm, PSBAX, to infer properties of unknown functions by selecting points based on posterior probability. Demonstrated its superior performance across multiple domains, including real-world protein engineering datasets. Contributed to a rigorous analysis of its convergence guarantees, advancing the theoretical foundations of posterior sampling. Currently collaborating with Lawrence Livermore National Lab to apply PSBAX to antibody portfolio optimization. This work is a finalist for the Undergraduate Operations Research Prize at the INFORMS 2024 Annual Meeting. **Paper:** Practical Bayesian Algorithm Execution via Posterior Sampling.

Machine Learning Researcher, Zhong Lab

2024

Princeton University

Summer Undergraduate Research Fellowship

- Research on investigating machine learning models for accelerating small molecule discovery by predicting chemical structures from NMR data with Prof. Ellen Zhong. Achieved 98.9% accuracy using a multilabel classifier to identify amino acid groups in peptides from synthetic NMR spectra. Ongoing work focuses on refining model architectures and expanding datasets for full sequence prediction. The work aims to eventually automate structure elucidation from new natural products, which has profound impact in drug development and chemical biology.

Machine Learning Researcher, Rigorous Systems Research Group

2022 - 2023

California Institute of Technology

- Research with Prof. Adam Wierman focused on optimization and control, particularly in the realm of online policy selection. Investigated the generalization of optimization algorithms to nonlinear dynamical systems with non-convex reward functions.

Research Assistant

2022

California Institute of Technology / Tsinghua University

Summer Undergraduate Research Fellowship

- Research with Prof. Yanan Sui on high-dimensional sampling using Bayesian optimization with Gaussian processes, specifically focused on its application in spinal stimulation. Employed Markov chain Monte Carlo techniques and Langevin dynamics to efficiently sample from intractable posteriors and devised an algorithm capable of maintaining an adaptive mesh on a high-dimensional action space. **Paper:** Improving Sample Efficiency of High Dimensional Bayesian Optimization with MCMC.

Intern, Tencent (Tencent Xinghuo Program)

2021 - 2022

Shenzhen, Tencent Quantum Lab

- Reproduced quantum approximate optimization algorithm using Tensorflow and Qiskit. Studied quantum information and quantum computation with focus on algorithms such as Grover search and variational algorithms.

TEACHING EXPERIENCES

Teaching Assistant

2022 - 2024

California Institute of Technology

- Responsible for leading recitations, holding weekly office hours, and writing up and grading problem sets. The average class size is 100 students.
- Courses: Data Mining and Machine Learning (2023 Winter & 2024 Winter), Statistical Inference (2023 Spring & 2024 Spring), Applied Linear Algebra (2024 Fall), Introduction to Probability Models (2023 Fall).

EXTRACURRICULAR EXPERIENCES

Competitions Expert, Kaggle

2021 - 2022

Online

- Sartorius - Cell Instance Segmentation: Top 4%, Rank 46/1505. Computer vision contest for object detection. Used popular frameworks such as Mask R-CNN and libraries such as MMDetection and Detectron2.
- Feedback Prize - Evaluating Student Writing: Top 8%, Rank 159/2058. NLP contest for text segmentation and classification. Used models such as Transformer, BERT, ELMo.

TECHNICAL SKILLS

Languages: Python, C, Java, OCaml, Swift, MATLAB.

ML Packages: PyTorch, TensorFlow, Pandas, SciPy.